



Machine Learning Introduction



Parcours Progis
Etudes, Medias, communication, Marketing
15.09.2026
Bahareh Afshinpour

Course Information

- **Bahareh Afshinpour**
- **Email:** bahareh.afshinpour@iepg.fr
- **Home page:** <http://afshinpour.com/>
- **Course time**
 - 1h30 TD
 - 1h30 TP

Textbook and References

- Ethem Alpaydin, "Introduction to Machine Learning", MIT Press, Prentice Hall of India, 3rd 2014.
- Neural Networks for Pattern Recognition by C. Bishop
- Python Machine Learning by S. Raschka and V. Mirjalili (highly recommended).
- YouTube videos.

Skills you will acquire

After completing this course, you will be able to:

- ✓ explain, compare, and contrast various machine learning topics and concepts like supervised learning, unsupervised learning, classification, regression, and clustering.
- ✓ You will also be able to describe how the various machine learning algorithms work.
- ✓ And finally, you will learn how to apply these machine learning algorithms in Python using various Python libraries.

Grading

- | | |
|--------------------------|-----|
| • TP (Homeworks- Report) | 20% |
| • Mid-Term Exams | 20% |
| • Final Exam | 40% |
| • Final project | 20% |

Policies

- Attendance is required.
- All submitted work must be yours.
- Protect your efforts! Don't let others see your codes, don't give others your results.
 - Lending your codes to others or allowing others to copy your work will be considered as collusion, thus receiving the same punishment as the plagiarist.
- Your codes need to be able to generate the results.

Computing needs

- **Basic Desktop or Laptop**
- **Ability to :**
 - **Install and run Python + notepad++**
 - » (Python offers a large number of packages that make it simple to get started and build applications of all complexity levels.)
 - **Jupyter Notebook**
 - » (The Jupyter Notebook is a simple interactive environment for programming with Python, which makes it really easy to share your results.)

Data in everyday life

- This is the age of “big data”

Each online interaction can generate data:

a click, a purchase, a comment, or time spent watching a video.

Organisations analyse these data to understand behaviour and make predictions.

Examples

- Marketing: predict whether a customer will respond to an offer.
- Social media: recommend content a user may find interesting.
- Opinion research: identify themes in survey responses.

- ✓ We want to have products and services specialized for us.
- ✓ We want our needs to be understood and interests to be **predicted**.
- ✓ There are certain **patterns** in the data.



#76651093

<https://www.lebigdata.fr/definition-big-data>

What is an algorithm?

- To solve a problem on a computer, we need an algorithm.



<https://www.scribbr.com/ai-tools/what-is-an-algorithm/>

An algorithm is a sequence of clearly defined steps for completing a task.

Example: sorting products by price

Input: €35, €10, €25

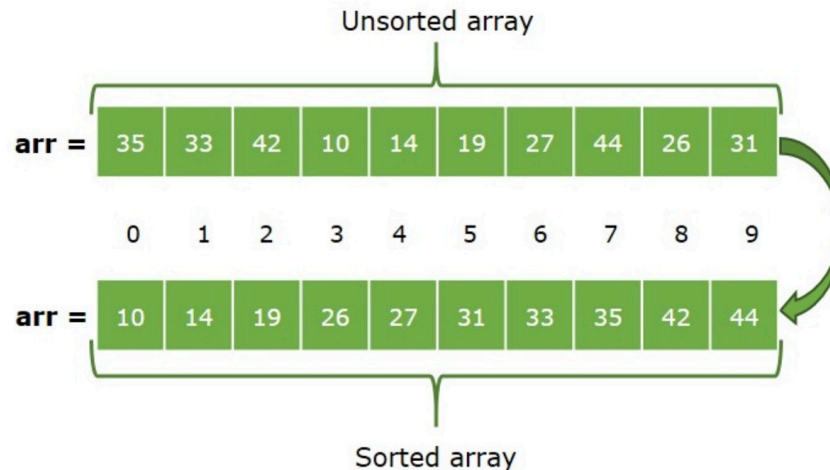
Steps: Select the lowest remaining price, add it to the ordered list, and repeat until no prices remain.

Output: €10, €25, €35

A programmer specifies the steps the computer follows.

What is an algorithm?

- For example, one can devise an algorithm for sorting. The input is a set of numbers and the output is their ordered list.



https://www.tutorialspoint.com/python/python_sort_arrays.htm

Sorting prices has clear rules we can write down.
But how would we write rules to predict whether someone will
like a video?

In machine learning, we use **examples** to learn patterns that help
make predictions.

Introduction- Algorithm

- **For some tasks, however, we do not have an algorithm.**
 - ✓ Predicting customer behavior is one;
 - ✓ another is to tell spam emails from legitimate ones.
 - what we want is to “**learn**” what constitutes spam from them.

In other words, we would like the computer(machine) to extract automatically the algorithm for this task.

Any piece of software that will consume training examples,
in order to make decisions over **unseen** data
without explicit programming is considered
learning.

Machine Learning application

Area	Example task
Marketing	Predict which customers are likely to cancel a subscription
Social media	Recommend content based on past interactions
Opinion research	Classify survey comments by topic or expressed sentiment
Management	Forecast product demand to help plan inventory
Finance	Flag unusual transactions for fraud investigation
Healthcare	Help identify suspicious areas in medical images

Choose one application. What data would the model need, and what would it predict or identify?

Machine Learning application

- **social media domain :**

Example: Facebook, Instagram, TikTok, and YouTube use ML to personalize users' feeds. Algorithms analyze your **past interactions** (likes, shares, watch time) **to predict and recommend** content you're likely to engage with.

- **political domain:**

Sentiment Analysis & Public Opinion Mining : ML models analyze social media posts, news articles, and surveys to gauge public sentiment toward policies, candidates, or political events. Tools like Brandwatch or Crimson Hexagon are used for real-time sentiment tracking.

What Is Machine Learning?

What is machine learning?

Machine learning uses algorithms that learn patterns from data to make predictions or identify structure.

Example: predicting customer cancellation

- **Experience:** past customer records, including whether customers cancelled.
- **Task:** predict which customers are likely to cancel.
- **Performance:** how well the predictions match outcomes for customers the model has not seen during training.

What Is Machine Learning?

The real question is what is learning?

Using past experiences to improve future performance.

For a machine, experiences come in the form of **data**.

What Is Machine Learning?

- A model is defined by **parameters**. Learning involves running a program to **optimize these parameters** using training data or past experience.
- The model may be **predictive** to make predictions in the future, or **descriptive** to gain knowledge from data, or bot.

Predicting customer purchases

A company plans to **send a promotional email**.

Goal: predict which customers will make a purchase within seven days of receiving it.

We have examples from previous campaigns, including customer information and whether each customer purchased.

What information about a customer might help us make this prediction?



Learning from past examples

Customer	Purchases in previous 6 months	Days since last purchase	Purchased within 7 days of email?
A	6	10	Yes
B	0	180	No
C	4	20	Yes
D	1	120	No
E	5	15	?

Features: information available before sending the email.

Label: the purchase outcome we know for past customers.

Prediction: the unknown outcome for another customer.

How the machine learns

Would you predict “Yes” or “No” for customer E? What influenced your answer?

You looked for patterns in the examples.

A learning algorithm also uses patterns, but we must test whether its predictions work on other customers.

Learning from past customer data

Training data

We use customer records from previous email campaigns.

Features

For each customer, we record information available before the email, such as previous purchases and days since the last purchase.

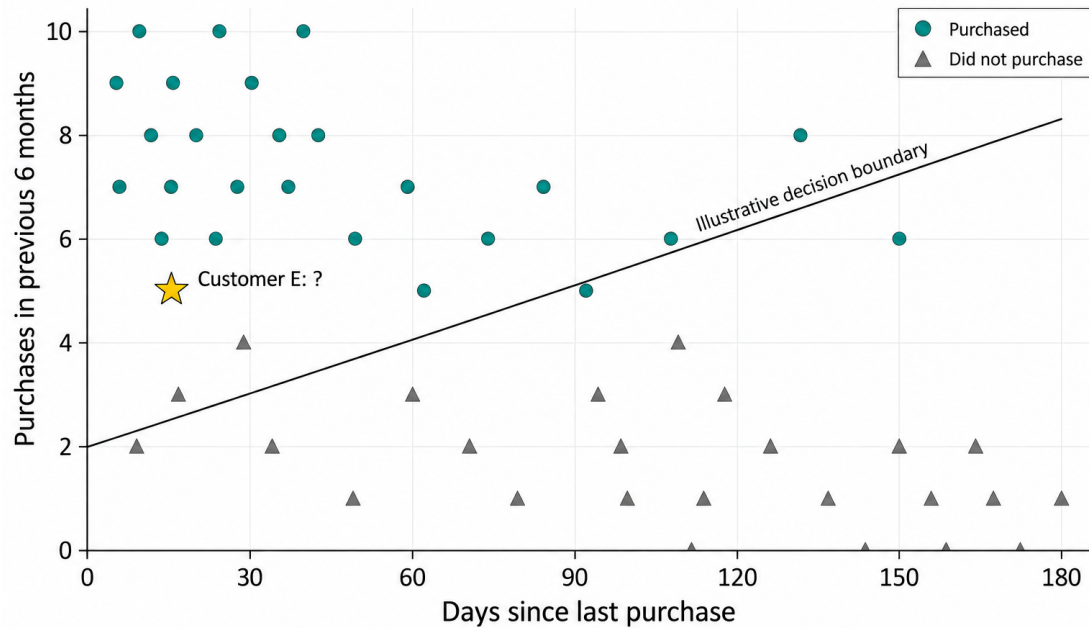
Labels

We know whether each customer purchased within seven days of receiving the email: **Yes or No**.

Learning and prediction

A learning algorithm uses these examples to fit a model. The model then predicts purchase likelihood for customers it did not see during training.

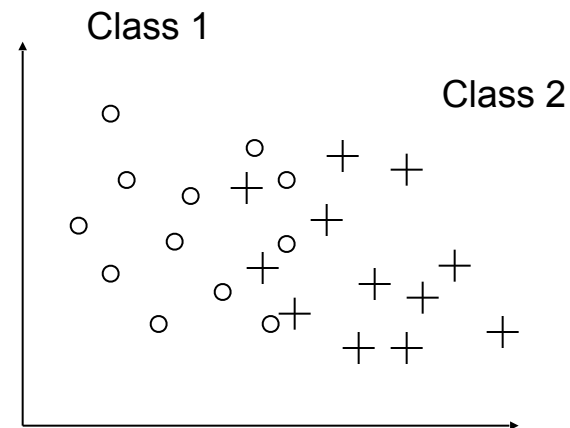
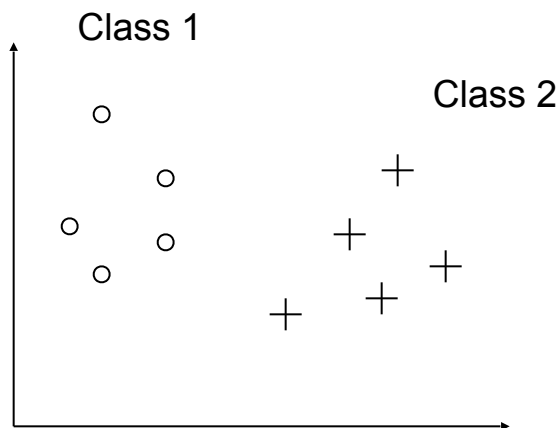
Predicting customer purchases



Fictional data. Purchase outcome within 7 days of the email.

Feature

- **Features** are numerically expressed properties of the signal.
- The set of features used for pattern recognition is called **feature vector**.
- The number of used features is the **dimensionality** of the feature vector.
- n-dimensional feature vectors can be represented as points in n-dimensional **feature space**.



Feature selection

- For example **having macrowave** is not very help ful feature for buying house



Useful for **efficiency**
of predictions and
interpretability

Lot size
Single Family
Year built
Last sold price
Last sale price/sqft
Finished sqft
Unfinished sqft
Finished basement sqft
floors
Flooring types
Parking type
Parking amount
Cooling
Heating
Exterior materials
Roof type
Structure style

All Features



Feature Selection



Final Features



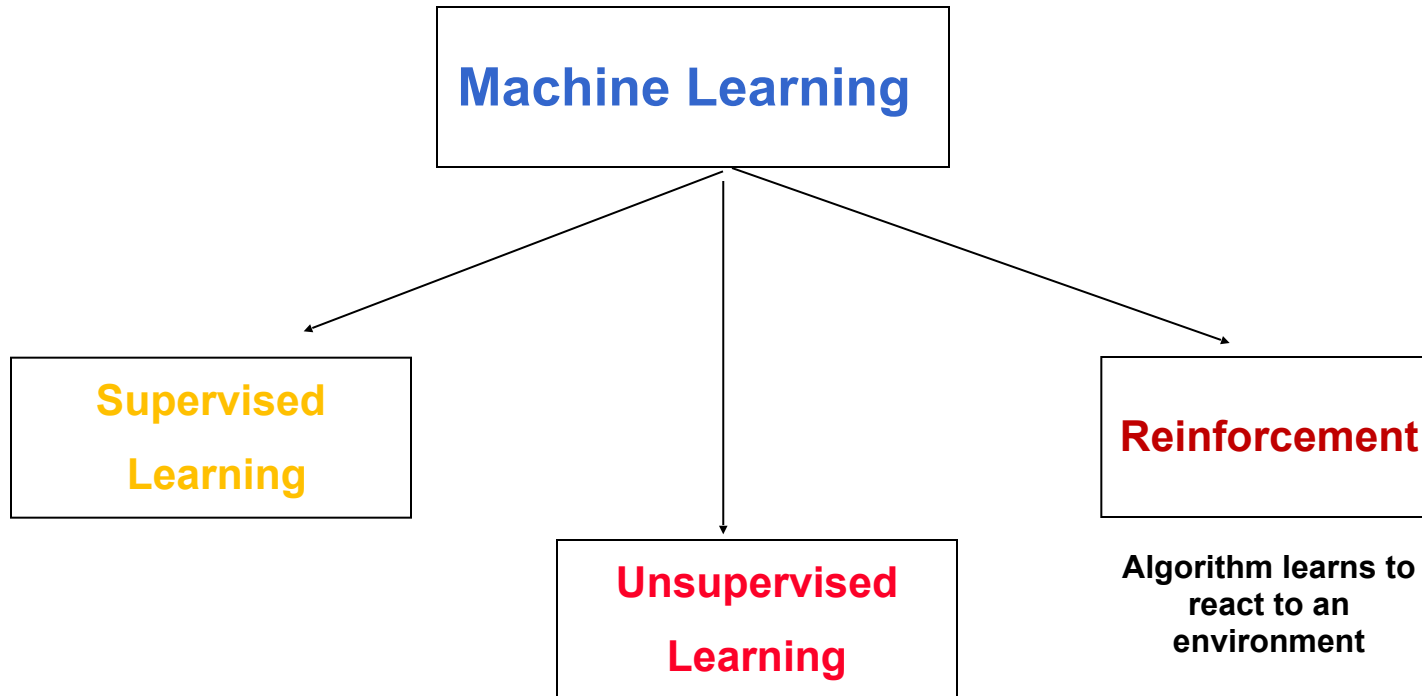
<https://vitalflux.com/machine-learning-feature-selection-feature-extraction/>

Feature extraction

But Feature **extraction** is about
extracting/deriving information from the original features set to create a
new features subspace.

-Dimensionality Reduction

About the course



- We'll cover the most widely used machine learning methods such as
 - ✓ Regression method
 - ✓ Classification methods
 - ✓ Clustering methods
 - ✓ Random forests
 - ✓ Neural nets, and so on.

Supervised Learning Problem

Supervised Learning Problem

How do we supervise a machine learning model?

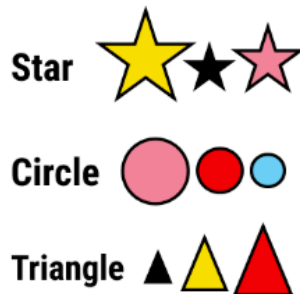
We do this by teaching the model

How exactly do we teach a model?

We teach the model by training it with some data from a labeled dataset

what does a labeled dataset look like?

It could look something like this



ID	Clump	UnifSize	UnifShape	MargAdh	SingEpiSize	BareNuc	BlandChrom	NormNucl	Mt	Class
1000025	5	1	1	1	2	1	3	1	1	benign
1002945	5	4	4	5	7	10	3	2	1	benign
1015425	3	1	1	1	2	2	3	1	1	malignant
1016277	6	8	8	1	3	4	3	7	1	benign
1017023	4	1	1	3	2	1	3	1	1	benign
1017122	8	10	10	8	7	10		7	1	malignant
1018099	1	1	1	1	2	10	3	1	1	benign
1018561	2	1	2	H	2	1	3	1	1	benign
1033078	2	1	1	1	2	1	1	1	5	benign
1033078	4	2	1	1	2	1	2	1	1	benign

Supervised Learning Problem

- In supervised learning,
 - we are given a data set and already know what our right output should be
 - with the assumption that there is a relationship between the input and the result
 - you have to study in relation with a different set of features called independent features.
- Two types of supervised learning tasks:
 - "Regression"
 - "classification"

Regression	Classification
Target: Continuous data	Target: Categorical data
Aim: Predict the value	Aim: Predict class/class probability

How does a supervised problem look?

X1	X2	X3	X4	X5	Y1	Y2
7	1	3	1	2	50	0
2	4	7	12	9	24	1
5	9	1	10	6	16	0
8	11	2	3	4	62	1
1	5	5	7	1	12	0

- x_1, x_2, x_3, x_4 are independent features
- Y_1 and Y_2 are dependent features
- Y_1 is continuous
- Y_2 is categorical

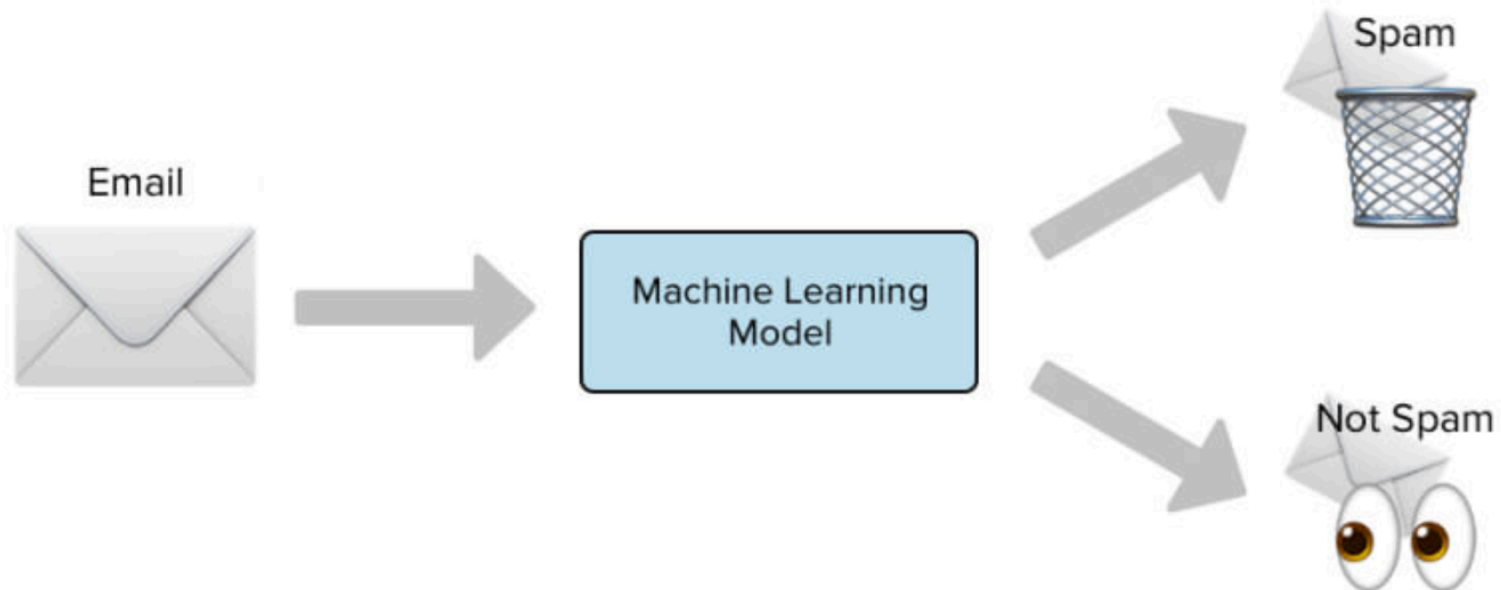
The outcome of a supervised learning problem is to have a dependent feature defined as a function of independent feature.

For example $Y_1 = f(X_1, X_2, X_3, X_4)$

How does a supervised problem look?

- **Input (Features):** Words, phrases, sender info, email length.
- **Output (Label):** "Spam" or "Not Spam".

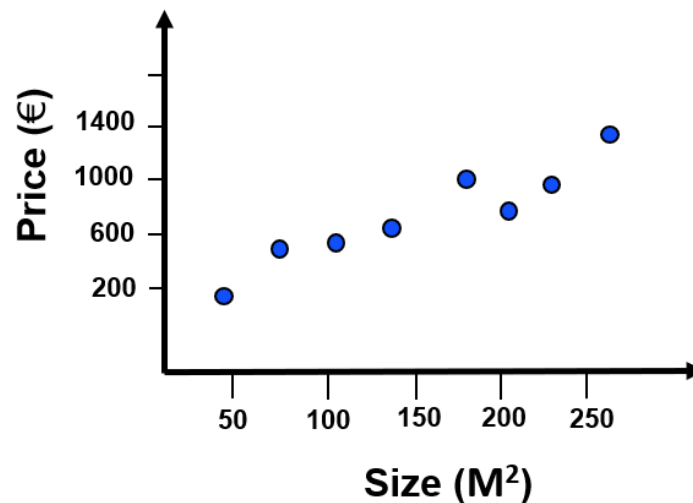
The dataset contains emails, each labeled as spam or not. The model learns patterns that distinguish spam from legitimate emails.



<https://www.superannotate.com/blog/supervised-learning-and-other-machine-learning-tasks>

Example of a Supervised Learning algorithm

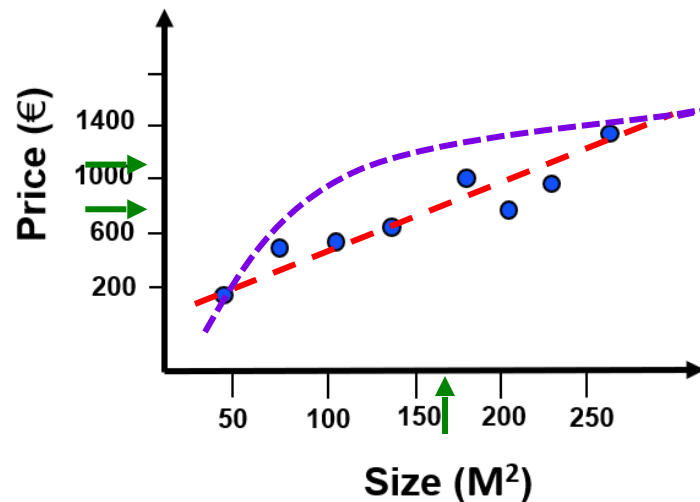
- Given data about the size of houses on the real estate market, try to predict their price.
- Price as a function of size is a **continuous** output, so this is a regression problem.



Imagine your friend has a 170 m² house and wants to know the ideal rental price. A learning algorithm can help by analyzing data from similar houses (size, location, features) and predicting a fair rental value based on past trends.

Example

- Put a straight line through the data, also fit a straight line to the data. It looks like maybe their house can be sold for about \$700. But maybe this isn't the only learning algorithm you can use, and there might be a better one.



- Instead of using a straight line, we fit a curve (like a quadratic) to the data—predicting the house could rent for about \$800.

What is regression?

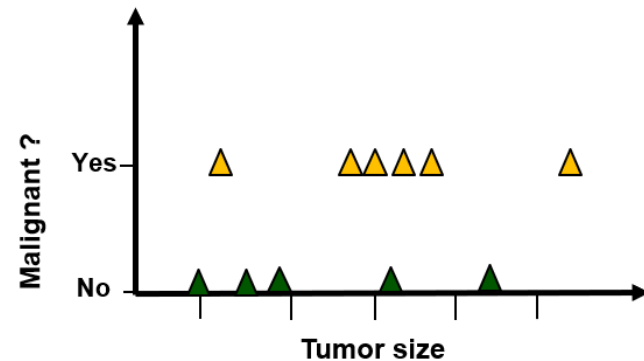
Regression is the process of predicting continuous values.

What is Classification?

Is the process of predicting discrete class labels or categories.

Examples

- Given a patient with a tumor, we have to predict whether the tumor is malignant or benign.
- a malignant tumor
 - is a tumor that is harmful and dangerous
- a benign tumor
 - is a tumor that is harmless



The Machine Learning question is, can you estimate what is the probability, what's the chance that tumor X as malignant versus benign?

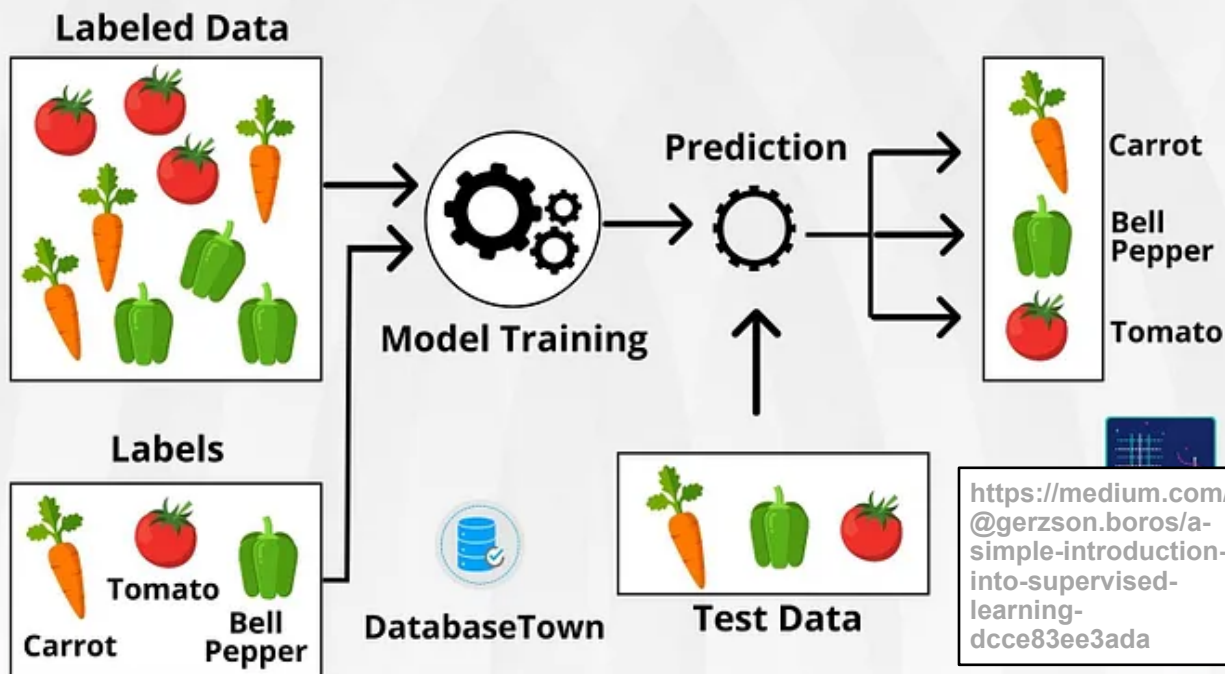
-
- For some learning problems what you really want is not to use like two or three features
 - For example, the age of the patient and the size of the tumor are two features
 - instead you want to use an infinite number of features

How do you deal with an infinite number of features?



SUPERVISED LEARNING

Supervised machine learning is a branch of artificial intelligence that focuses on training models to make predictions or decisions based on labeled training data.



TP1

-
- Introduction into *Python* programming with:
 - The traditional "hello world";
 - Some data structures and variables.
 - Datatypes
 - If

WHY PYTHON

Readable code: we can start with short, simple programs.

Free to use: students can practise outside class.

Tools for data analysis: explore customer records and survey responses.

Tools for machine learning: build models and evaluate predictions.

Repeatable analyses: rerun the same steps when new data arrive.



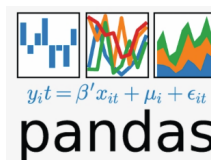
SO HOW DO I GET STARTED?



PYTHON LIBRARIES FOR DATA SCIENCE

Library	What we use it for
pandas	Load, clean, and explore customer or survey data
NumPy	Work with numerical arrays and calculations
Matplotlib / seaborn	Create charts and visualise patterns
scikit-learn	Train models and evaluate their predictions

matplotlib



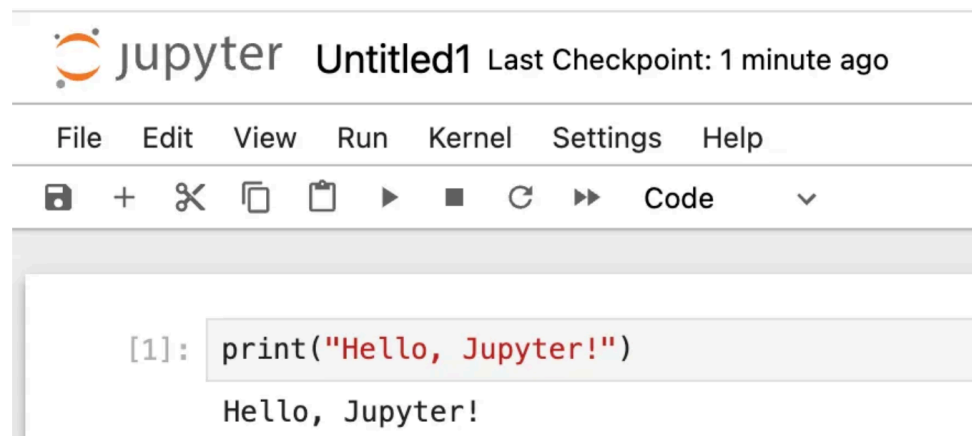
PYTHON SCIKIT-LEARN



- **Popular machine learning toolkit in Python** <http://scikit-learn.org/stable/>
 - Good documentation
 - Is really easy to implement
 - Has most of the classification, regression and clustering algorithm
- **Requirements**
 - Anaconda
 - Available from <https://www.continuum.io/downloads>
 - Includes numpy, scipy, and scikit-learn (former two are necessary for scikit-learn)

JUPYTER

- Jupyter is a freely available web application
- Jupyter promotes collaboration and reproducibility by allowing users to share their notebooks with others via email, GitHub, or the Jupyter Notebook Viewer.



PICK UP GOOD HABITS RIGHT AWAY!

- Comments in your code help you or someone else understand
 - What your program does
 - What a particular line or section of code does
 - Why you chose to do something a particular way
 - Anything that might be helpful to know if I am looking at the code later and trying to understand it!

IN PYTHON WE USE A # TO INDICATE COMMENTS

```
#My first Python Application  
#Created by me!  
#Print command displays a message on the screen  
print('Hello World')
```

Did you notice the colors?

ERRORS IN PYTHON

- It is important to read error messages carefully.

```
# Print string as error message
```

```
frint("Hello, Python!")
```

```
-----  
NameError                                Traceback (most recent call last)  
Cell In[2], line 3  
    1 # Print string as error message  
----> 3 frint("Hello, Python!")  
  
NameError: name 'frint' is not defined
```

```
# Try to see built-in error message
```

```
print("Hello, Python!")
```

```
Cell In[3], line 3  
    print("Hello, Python!")  
      ^  
SyntaxError: unterminated string literal (detected at line 3)
```

Python interprets
your script line by
line as it executes it.
Python will stop
executing the entire
program when it
encounters an error

DATA TYPES

- A type is how Python represents different types of data.
 - Integers, real numbers, string and boolean
- In Python, a string is a sequence of characters.
 - A string can be spaces or digits. A string can also be special characters.
- You can change the type of the expression in Python, this is called typecasting.

```
# Integer
```

```
11
```

```
# Float
```

```
2.14
```

```
# String
```

```
"Hello, Python 101!"
```

VARIABLES

- **We can use variables to store values.**
- **We assign a value to a variable using the assignment operator, i.e, the equal sign.**
- **We can then use the value somewhere else in the code by typing the exact name of the variable.**
- **We can use the type command in variables as well.**

PYTHON CONDITIONS AND IF STATEMENTS

Python supports the usual logical conditions from mathematics:

- Equals: **a == b**
- Not Equals: **a != b**
- Less than: **a < b**
- Less than or equal to: **a <= b**
- Greater than: **a > b**
- Greater than or equal to: **a >= b**

```
a = 33
b = 200
if b > a:
    print("b is greater than a")
```

b is greater than a

An "if statement" is written by using the **if** keyword.

ELIF AND ELSE

- The **elif** keyword is python's way of saying "if the previous conditions were not true, then try this condition".

```
a = 200
b = 33
if b > a:
    print("b is greater than a")
elif a == b:
    print("a and b are equal")
else:
    print("a is greater than b")
```

a is greater than b

In this example a is equal to b, so the first condition is not true, but the elif condition is true, so we print to screen that "a and b are equal".

LOADING PYTHON LIBRARIES

```
In [ ]: #Import Python Libraries  
import numpy as np  
import scipy as sp  
import pandas as pd  
import matplotlib as mpl  
import seaborn as sns
```

Press Shift+Enter to execute the *jupyter* cell

Helpful Web Site in Python

The Deck is Stacked Against Developers

Machine learning is taught by academics, for academics.

That's why most material is so *dry* and *math-heavy*.

Developers need to know what works and how to use it.

We need *less math* and *more tutorials with working code*.



Welcome to Machine Learning Mastery!

Hi, I'm Jason Brownlee PhD and I help developers like you skip years ahead.

Discover how to get *better results, faster*.

Click the button below to get my free EBook and accelerate your next project
(and access to my exclusive email course).

[Send Me the Free eBook!](#)

Join over 150,000 practitioners who already have a head start.

END

Occam's razor

simpler explanations are more plausible and any unnecessary complexity should be shaved off